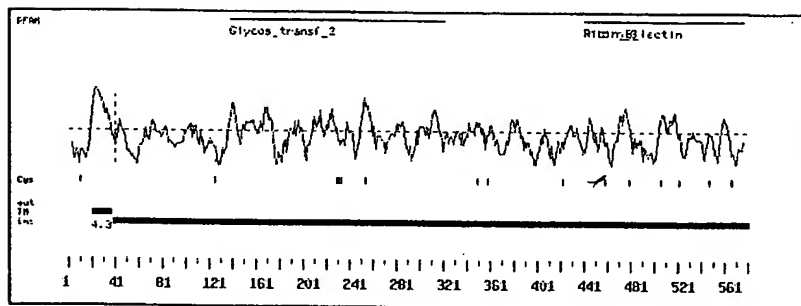


[Back to refseq.cgi](#)

Analysis of 33945 (581 aa)



>33945
MMRTAARPCPRELARGREALLVLLALLAAGLSVLRAGRGAGAGAAEPFPPTPRGR
REPVPFPVPANALGARGAARVLRQGEELRLQESVLRHQINILSDRI SLARLPXR
WNPFLCKEKYDYINLPRTSVIIAPYEAWSYLLRTVYVLETSFDILLEEVILVDDYSR
EHLKERLANELSGLPKVRILIRANKREGVLRARLGASAARGDVLTFDCHCEHEGMLP
LLQRIHEESAVVCPVLDVDMNTPEYLGNSGEPQIGGFDMRLVPTMTVPFERIRHQS
PVDVIRSPTHAGLFAVSKYFETLGSYDTGHEVWOGNLEFSFRWQGGVLETHPCSH
VGRVFRQAPFSRRLALANGVAEEVWHEPSEKELTHGIPARLEFPDPTERQIARLKL
CKKDFMTLETVPYELARVPEARPGPFQSGKQELTDYCFDTPPDQIVGRQVILTLCH
GQZQMFFETTSQKEIRYNTYHQPEQIAVEAGDYLINSLCEETAPEDQKFILOEKSLF
HEQSKKCVGAANKESDSFVPLKDCSTSDHQKMPFEREL

PSORT Prediction of Protein Localization

MITDISC: discrimination of mitochondrial targeting seq

R content: 8 Hyd Moment(75): 13.29
Hyd Moment(95): 16.54 G content: 1
D/E content: 2 S/T content: 1
Score: 2.30

Gavel: prediction of cleavage sites for mitochondrial preseq

R-2 motif at 19 RRC|PR

NUCDISC: discrimination of nuclear localization signals

pat4: none
pat7: PLCKEKK (3) at 123
bipartite: none
content of basic residues: 12.94
NLS Score: -0.22

ER Membrane Retention Signals:

XXRR-like motif in the N-terminus: MMRT

KKXX-like motif in the C-terminus: KERM

Final Results (k = 9/23):

43.5 %: mitochondrial
30.4 %: cytoplasmic
8.7 %: vacuolar
8.7 %: nuclear
4.3 %: Golgi
4.3 %: endoplasmic reticulum

prediction for 33945 in mit (k=23)

Start	End	Feature	Seq
86	107	Leucine zipper pattern (PS00029)	LQGEELRLQE...VRLHQINIVL

Signal Peptide Predictions for 33945

Method	Predict	Score	Mat@
SignalP (eukaryote)	YES		40

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
20	37	out->ins	4.3

>33945
MMRTAARPCPRELARGREALLVLLALLAAGLSVLRAGRGAGAGAAEPFPPTPRGR
REPVPFPVPANALGARGAARVLRQGEELRLQESVLRHQINILSDRI SLARLPXR
WNPFLCKEKYDYINLPRTSVIIAPYEAWSYLLRTVYVLETSFDILLEEVILVDDYSR
EHLKERLANELSGLPKVRILIRANKREGVLRARLGASAARGDVLTFDCHCEHEGMLP
LLQRIHEESAVVCPVLDVDMNTPEYLGNSGEPQIGGFDMRLVPTMTVPFERIRHQS
PVDVIRSPTHAGLFAVSKYFETLGSYDTGHEVWOGNLEFSFRWQGGVLETHPCSH

<http://img.mpl.com/seqanal/refseq/33945.12907.htm>

VGRVFRKQAPTSTKALANSVAAEVMDDEFKELYTHRNPRARLEPPGCVTERKQLADKL
 QCKDFRMFLTVYFELHVPEDRPGFQHLQHGKLTDTYCFDYNPPDSNQIVGHVILYLCH
 QHQQHFFFTTSQKEIRVTRHQPDQCIANENHDTLIHSLCEETAPEMOKPILQEDGSLF
 HQSKKCYQAARKESDSFVPLRACCTNSDQKMFTRDRL

Transmembrane segments for presumed mature peptide

Start	End	Orient	Score
-------	-----	--------	-------

Prosite Pattern Matches for 33945

Prosite version: Release 12.2 of February 1995

>PS00044|PDOC00004|CAMP_PHOSPHO_SITE CAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 552 RKES 555

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 5 TAR 7

Query: 55 TPR 57

Query: 97 SVR 99

Query: 108 SDR 110

Query: 178 SDR 180

Query: 318 SKK 320

Query: 343 SFR 345

Query: 411 TER 413

Query: 492 SQK 494

Query: 544 SKK 546

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 158 SVLE 161

Query: 162 TSPD 165

Query: 178 SDRE 181

Query: 225 TFLD 228

Query: 289 TVPE 292

Query: 300 SPVD 303

Query: 330 TGHE 333

Query: 492 SQKE 495

Query: 524 TAPE 527

Query: 567 TNSD 570

>PS00007|PDOC00007|TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site.

Query: 426 RMFLETVY 433

>PS00023|PDOC00008|MYRISTYL N-myristoylation site.

Query: 42 GAGAGA 47

Query: 207 GLVRAR 212

Query: 215 GASAAR 220

Query: 312 GGLFAV 317

Query: 326 GSYDTG 331

Query: 351 GVLETH 356

Query: 447 GHLQNK 452

Query: 481 GHQHQ 486

>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 58 PQRR 61

>PS00016|PDOC00016|RGD Cell attachment sequence.

Query: 220 RGD 222

>PS00022|PDOC00029|LEUCINE_ZIPPER Leucine zipper pattern.

Query: 86 LQGEELRLQEEVRLHQIHYL 107

Protein Family / Domain Matches, HMMer version 2

```
Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - -
HMM file: /prod/dm/seqanal/PFAM/pfam5.5/Pfam
Sequence file: /prod/dm/wspace/orfanal/oa-script.13910.seq
Query: 33945

Scores for sequence family classification (score includes all domains):
Model Description Score E-value N
-----
Glycosyl_transf_2 Glycosyl transferases 85.1 1.4e-21 1
Ricin_B_lectin Similarity to lectin domain of ricin 18.7 0.0016 1

Parsed for domains:
Model Domain seq-f seq-t hm-f hm-t score E-value
-----
Glycosyl_transf_2 1/1 139 322 .. 1 187 [] 85.1 1.4e-21
Ricin_B_lectin 1/1 441 577 .. 1 145 [] 18.7 0.0016

Alignments of top-scoring domains:
Glycosyl_transf_2: domain 1 of 1, from 139 to 322: score 85.1, E = 1.4e-21
*-slvtpcmdeadykellelvlaqg.tledieifivddgSetdette
s+if Yne+...L+++ Svl+ s++ + s+i vdd S+ + s+
33945 139 SVIIFYNEMWSTLARTVSVLETSdILLLEVLVDDYSRHLKE 185
leedylderikeenprilivirleemsgQdaaArnkqirratGdsdyllf
+ + + + + p+ +sif + + G +sR +g +s+Gd +s+f
33945 186 RLANEL-----SGLPK-VRLIRANKRE-GLVRLRLGASARGD--VLTF 226
LDaDdfitpdkleklidyeeatdaavvipaids....yeyaeasnlly.
LD+...+ +le+)+... + avv+ id+ + +s+ + + +
33945 227 LDCHDCHDGNLEFLQRIEESAVCVFIDVdwnPEYLGHSGEPQI 276
riaradterslfagllrktrltgglelsfeigsnllyrreafeelfc-+
++++ + + + + + E+ s+ + + + + + + + + + + + +
33945 277 GGFDMRLVFTMTVFERERIRHQP-VDIVRSPTHAGGLFAVSKTYF 322
Ricin_B_lectin: domain 1 of 1, from 441 to 577: score 18.7, E = 0.0016
*-rsvitvitil.vngmsgrClDvnssesdd...nqvqlvnChsrpknQ
+ + + + + l+ g + + C D n +s+ + qv l Ch+ +g nQ
33945 441 DRPGFPHLAgGLTDTCTFTHPPDDEQIVGRVILYLCHG-HGQNG 486
kseltYdesdqeirs.vvnnkClitvnanspseVkllycdsatachQd
+ + + + + s+ sif + + + C +v+ + + + + C + + +nQ+
33945 487 FFETV---SQKEIRYTHQPEGCIAGEAGHET--LIMHLCEETAPENQKF 531
elnndglignklllnlmtglvldvsgdsngTklilytcsggrnqgnl
l dg+ + + + + + + + + + + + + + + + + + +
33945 532 ILQEDGSLFH-----QSKKCVQAAKESdsfVPLLRDCTNSDHQNF 575
ptc-+
+
33945 576 FK 577

//
Searching for complete domains in SMART
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - -
HMM file: /ddm/robison/smart/smart/smart.all.hmms
Sequence file: /prod/dm/wspace/orfanal/oa-script.13910.seq
Query: 33945






Scores for sequence family classification (score includes all domains):
Model Description Score E-value N
-----
ricin_3 73.3 5.1e-18 1

Parsed for domains:
Model Domain seq-f seq-t hm-f hm-t score E-value
-----
ricin_3 1/1 445 577 .. 1 135 [] 73.3 5.1e-18

Alignments of top-scoring domains:
ricin_3: domain 1 of 1, from 445 to 577: score 73.3, E = 5.1e-18
*-rpyflilgntglCLdvnssesksdGnpvqlwdChg.ggnQlMklt
+ + + + + g t+ C don +s+ C +v+ + Chg g nQ + + t
33945 445 PFPGDQNGLLTYCTFTHPPDDEQIVGRVILYLCHGAGNQNFFETV 491
Ynesdgairi...nsdCLctmg...tvtlysCdgtldgndnQkWevnd
s+ +sif + + + + + C + + + + + C+ t + +nQ+ + + d
33945 492 ---SQKEIRYTHQPEGCIAGEAGHET--LIMHLCEET--APENQKFILQED 536
qtirmpknskkgvdsplclchvkdgn....kvqlwtcngsdapnQkMife
g + + + + + + + + + + + + + + + + + + + + + + +
33945 537 GSLPHE-----QSKKCVQAAKESdsfVPLLRDCTNS--DHQNF 577
c-+

33945 - - - - -
//
```

ProDom Matches

Prodomid	Start	End	Description	Score
View Prodom 2681 	101	138	p99.2 (20) PAGT(3) // TRANSFERASE N-ACETYL GALACTOSAMINYLTRANSFERASE POLYPEPTIDE ACETYL GALACTOSAMINYLTRANSFERASE UDP-GALNAC:POLYPEPTIDE GLYCOSYLTRANSFERASE PROTEIN-UDP PROTEIN- UDP N-	96
View Prodom 137344 	108	187	p99.2 (1) O44164_CAEEL // COSMID F16B3	72
View Prodom 93 	139	285	p99.2 (314) NODC(8) PAGT(3) CAG2(3) // PROTEIN TRANSFERASE GLYCOSYLTRANSFERASE PUTATIVE BIOSYNTHESIS GLYCOSYL SYNTHASE TRANSMEMBRANE N-ACETYL GALACTOSAMINYLTRANSFERASE MEMBRANE	108
View Prodom 2308 	287	443	p99.2 (23) PAGT(3) // N-ACETYL GALACTOSAMINYLTRANSFERASE TRANSFERASE POLYPEPTIDE ACETYL GALACTOSAMINYLTRANSFERASE UDP-GALNAC:POLYPEPTIDE GLYCOSYLTRANSFERASE PROTEIN-UDP PROTEIN- UDP N-	567
View Prodom 617 	453	574	p99.2 (70) PAGT(3) ABFB(2) E13B(2) // SIGNAL HYDROLASE PRECURSOR TOXIN RRNA N-GLYCOSIDASE GLYCOSIDASE LECTIN REPEAT GLYCOPROTEIN	104
Prodomid	Start	End	Description	Score

View Prodom 2308   

>2308 p99.2 (23) PAGT(3) // N-ACETYL GALACTOSAMINYLTRANSFERASE TRANSFERASE
POLYPEPTIDE ACETYL GALACTOSAMINYLTRANSFERASE UDP-GALNAC:POLYPEPTIDE
GLYCOSYLTRANSFERASE PROTEIN-UDP PROTEIN- UDP N-
Length = 172

Score = 567 (204.7 bits), Expect = 2.6e-55, P = 2.6e-55
Identities = 102/166 (61%), Positives = 128/166 (77%)

Query: 287 MHTVPERERIRMQS-PVDVIRSPHAGGLFAVSKKYFYLGSYTGTGHEWOGENLEFSFR 145
NW VFE ER R + P D IASPHAGGLFA+K+YFE LG+YD GM+MOGENLE SFR
Sbjct: 2 MHTVPEERKRRKPTDPIASPHAGGLFAINKEYFEELGYDQNDIMOGENLEFSFR 61
Query: 346 INQCGVLETHPCSHVGVFRKAPYS-----RKALANSVKAEEVWDEFELTYHR 398
+MQCG LE PCSHVGVFRK+FT+ ** N+V AEVWD+KE +T
Sbjct: 62 VMQCGRLIVPCSHVGVFRKASFTYFGKDSGKDVISNIVRVAEVMWDTKEYFYKH 121
Query: 399 NFRAR-LEPPGVTERQRLKRLQCKDFQFLETYYFELRVPEDP 443
NP+AR + PGD+ERK+LR+KLGCK FDM+LE VYF+L+VF P
Sbjct: 122 NPQARKVRDFDISERKRLKRLQCKSFQWLENVYDOLYVPAHEP 167

View Prodom 617   

>617 p99.2 (70) PAGT(3) ABFB(2) E13B(2) // SIGNAL HYDROLASE PRECURSOR TOXIN
RRNA N-GLYCOSIDASE GLYCOSIDASE LECTIN REPEAT GLYCOPROTEIN
Length = 140

Score = 104 (41.7 bits), Expect = 3.1e-05, P = 3.1e-05
Identities = 29/128 (22%), Positives = 55/128 (42%)

Query: 453 GLTDYCTDYNPDENQIVGHVILYLCHQH-QNQFFETTSQKEIRYNTHQFSCIAVEA 511
G C D N + + G+ V L+ CH G M O + + + IR N + + + C+
Sbjct: 11 GNGMCLDNDGANTVLGQNFVGLQCHNMGKQWMTFMTDITRSHKDKR-CLTSSG 69
Query: 512 ---GHDTLIMLCETAPENQKFLQEDGSLFHE-OSKCYQAARKESDSFVPLLRDC 566
G +I + MQK+ DG+ S C+ A + + + + + C
Sbjct: 70 YSPVGSSVHITNCHSARKGAKWJWFNSDGTIIQNPMSHCLDAVNSDTANGTKIMVTC 129
Query: 567 TNSNQKN 574
+ Q+N
Sbjct: 130 NGNSNQKN 137

View Prodom 2681

>2681 p99.2 (20) PAGT(3) // TRANSFERASE N-ACETYL GALACTOSAMINYLTRANSFERASE
POLYPEPTIDE ACETYL GALACTOSAMINYLTRANSFERASE UDP-GALNAC:POLYPEPTIDE
GLYCOSYLTRANSFERASE PROTEIN-UDP PROTEIN- UDP N-
Length = 118

Score = 96 (38.9 bits), Expect = 0.00022, P = 0.00022
Identities = 21/39 (53%), Positives = 26/39 (66%)

Query: 101 HQINILYLSRISLHRLFXNNPLCKKKYD-YNNLPRT 138
+Q N+Y SD ISL+R LP CK +RT+ YNNLP T
Sbjct: 80 MQFLYASINSLSRLSPVPRPECKTKTKNPYINLPTT 118

View Prodom 93

>93 p99.2 (314) NODC(8) PAGT(3) CAG2(3) // PROTEIN TRANSFERASE
GLYCOSYLTRANSFERASE PUTATIVE BIOSYNTHESIS GLYCOSYL SYNTHASE TRANSMEMBRANE
N-ACETYL GALACTOSAMINYLTRANSFERASE MEMBRANE
Length = 196

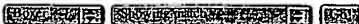
Score = 108 (43.1 bits), Expect = 0.00056, P = 0.00056
 Identities = 46/163 (28%), Positives = 74/163 (45%)

Query: 139 SVIIAPYNEAMSTLRTVYSVLETSPO--ILAEVILVDYS-DR--EHLKE-ALANELS 192
 SVII YNE S L V SVL* * *EE+I+VDD S D E +E + E
 Sbjct: 3 SVIIPYNEELSILATLVSSVLQQYENHYDGEIIIVDGGSTANTAEIVEEYSYSAEKE 62

Query: 193 GLPKVRLIRANKREXXXXX-----XXXXXXXXXXDLTFLDCHCCHED--WLEPLAQ 243
 KV+IR K D + FLD + H WLE L+*
 Sbjct: 63 SRVKKVIRNEDKNSCHSSAMKGLKHALSCAGDIIIVFLDSD-DVDMSPOMLEKLE 121

Query: 244 RIHEESAVVCVVIDVIMNTFEYLGNSGEPQI-GGPDWLVF 285
 + E + VV V+ + +Y + * GG+W +*+
 Sbjct: 122 AMEDNADVVVGSQVVDENKQYSSATRLINEDGTHWMMY 164

View Prodom 137344



>137344 p99.2 (1) 044164_CAEEL // COSMID F1693
 Length = 196

Score = 72 (30.4 bits), Expect = 8.3, P = 1.0
 Identities = 23/80 (28%), Positives = 40/80 (50%)

Query: 108 SDRISLHRLPKRWNPFLCKEXYDTMLPRISVIIAPYNEAMSTLRTVYSVLETSPOIL 167
 S + R P R PL K + D D L RTS + * NS-L + + +* DI+
 Sbjct: 8 SSVKSPRSKPKR-PFLAKRQTVDEALSRISWGNKKDKREHSSL-SNITKSKFDII 65

Query: 168 LEEVILVDYSOREHLAKRL 187
 E + +*+ + R L ER+
 Sbjct: 66 GERLTIIIDQINSRLALLERV 85

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.